DUPLICATE ALSO

Forecasting Research

Forecasting Research
Scientific Paper No. 50

Applications of the Monge-Ampère equation and Monge transport problem to meteorology and oceanography

by

M J P Cullen and R J Douglas

May 1998

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

ORGS UKMO F

National Meteorological Library
FitzRoy Road, Exeter, Devon. EX1 3PB

The Met.Office

Forecasting Research
Scientific Paper No. 50

Application of the Monge-Ampère equation and Monge transport problem to meteorology and oceanography

by

M.J.P. Cullen and R.J. Douglas

May 1998

This will appear in the proceedings of the NSF-CBMS Conference on the Monge-Ampere equation - applications to geometry and optimisation, Boca Raton, Florida, July 1997.

Numerical Weather Prediction
Room 344
Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

© Crown Copyright 1998

Permission to quote from this paper should be obtained from the above Meteorological Office division.

Please notify us if you change your address or no longer wish to receive these publications.

Applications of the Monge-Ampère equation and Monge transport problem to meteorology and oceanography

M.J.P. Cullen
Forecasting Research Division,
Meteorological Office,
London Road,
Bracknell, Berkshire RG12 2SZ, U.K.

and R.J. Douglas
Department of Mathematics,
University of Reading,
Whiteknights, P.O. Box 220,
Reading RG6 6AX, U.K.

Abstract

The Monge-Ampère equation and Monge mass transport problem are shown to arise naturally in a system of equations which describes large scale motions of the atmosphere and ocean. In the case of constant rotation, it is shown that solutions of this system of equations can be interpreted as a sequence of minimum energy states: the energy minimisation can be rewritten as a Monge mass transport problem. Given a technical assumption, we can identify the solution of this problem with the solution of a Monge-Ampère equation: we can rewrite the system of equations as a coupled Monge-Ampère/transport problem. In the general case where the rotation is a function of horizontal position, we use a recent theorem of McCann to identify a transformation which corresponds to that defined for the constant rotation case.

1 Introduction

This paper studies the semigeostrophic equations, a standard model for slowly varying flows constrained by rotation and stratification. They have, in particular, been used to study front formation in meteorology. We show that Monge mass transport problems and Monge-Ampère equations arise naturally when we seek solutions of these equations.

The paper is organised as follows. In section 2, we show how the semigeostrophic equations can be obtained by approximating the general equations of motion for atmospheric and oceanic flow. For constant rotation, we

introduce the geostrophic transformation $\mathbf{X}$: from these values we can recover the velocity, potential temperature, and pressure of the fluid. In section 3 we demonstrate that (stable) solutions of the semigeostrophic equations can be interpreted as a sequence of minimum energy states. We show that the following procedure is well posed and yields the same answer as solving the semigeostrophic equations: at each fixed time t we predict the values of $\mathbf{X}$ on particles from an evolution equation for $\mathbf{X}$, and then arrange the particles to minimise geostrophic energy. The minimisation problem is shown to be equivalent to a Monge mass transport problem: in section 3.2 we obtain that at each fixed time t, $\mathbf{X}$ is equal to the gradient of a convex function. Moreover if $\mathbf{X}(t,.)$ is nondegenerate in the sense of Brenier [**3**] for each time t, we show in section 3.3 that we can find the mapping which relates the Lagrangian and Eulerian variables. (The concept of a nondegenerate function is recalled in section 3.2.)

Retaining the nondegeneracy assumption, in section 4.1 we rewrite the semigeostrophic equations as a coupled Monge-Ampère/transport problem. Section 3 demonstrated that $\mathbf{X}$ is equal to the gradient of a convex function at each time t: we now show that the Legendre-Fenchel transform (or conjugate convex function) of the convex function satisfies a Monge-Ampère problem in a generalised sense. The evolution equation is rewritten as a transport equation: the advecting velocity may be obtained from the Monge-Ampère problem. We discuss existence and uniqueness of solution for the coupled system with reference to relevant results in the literature. Section 4.2 considers properties of the long time evolution of the semigeostrophic equations. In section 5, we study semigeostrophic theory where the rotation is a function of horizontal position. We use an energy minimisation problem to define a transformation that corresponds to the geostrophic transformation for constant rotation. This problem is shown to be equivalent to a Monge mass transport problem on a Riemannian manifold, and its solution is due to a recent theorem of McCann [**17**]. These results are stated in section 5.2.

2 Equations for the large scale evolution of the atmosphere and ocean

In this section we state the general equations of motion for atmospheric and oceanic flow, and show how the semigeostrophic equations can be obtained by approximating the momentum in the general equations by its geostrophic values. For constant Coriolis parameter f the geostrophic transformation is introduced, and we study the semigeostrophic equations in these new variables.

2.1 General equations of motion

We start with an incompressible Boussinesq form of the 3-dimensional Euler equations for a rotating stratified fluid.

$$\begin{aligned} \frac{Du}{Dt} + \frac{\partial p}{\partial x} - fv &= 0 \\ \frac{Dv}{Dt} + \frac{\partial p}{\partial y} + fu &= 0 \\ \frac{Dw}{Dt} + \frac{\partial p}{\partial z} + g\theta/\theta_0 &= 0 \\ \frac{D\theta}{Dt} &= 0 \\ \nabla.(u,v,w) &= 0 \end{aligned} \tag{1}$$

In these equations (x, y, z) are Cartesian coordinates, (u, v, w) is the velocity field and $\frac{D}{Dt}$ is the Lagrangian derivative. p is the pressure. The system is assumed to be rotating with angular velocity $f/2$, which may be a function of horizontal position (x, y). θ is the potential temperature (in the atmospheric case) which plays the role of density under the Boussinesq approximation. In the oceanic case density is used directly. θ_0 is a reference value of θ. g is the acceleration due to gravity.

The derivation of these equations in the atmosphere relies on the fact that the main effect of compressibility is expressed as a strong vertical pressure gradient in hydrostatic balance. It can then be shown, e.g. Holton [**12**], that equations of the form (1) can be derived by assuming z is a function of pressure and making a form of Boussinesq approximation. This specific form of equation is derived in Hoskins [**13**]. In the ocean, (1) can be reached more naturally by simplifying the equation of state, which relates density to temperature and salinity.

2.2 The semigeostrophic equations

Weather systems and equivalent large scale flows in the ocean can be characterised as slowly varying flows constrained by rotation and stratification. In this asymptotic regime, the largest terms in the momentum equations of (1) express geostrophic and hydrostatic balance:

$$\begin{aligned} \frac{\partial p}{\partial x} - fv &= 0 \\ \frac{\partial p}{\partial y} + fu &= 0 \\ \frac{\partial p}{\partial z} + g\theta/\theta_0 &= 0 \end{aligned} \tag{2}$$

Equations (2) give no information about the time evolution. A standard approximate model for predicting the time evolution in this regime is given by approximating the momentum by its geostrophic value in (1). This gives the semigeostrophic system used by Hoskins [**13**]. This system is valid if the timescale for a change in velocity following a fluid particle is much greater than f^{-1}. In particular this is true in the case of front formation in meteorology, where the flow is largely parallel to the front. We state the equations in a form similar to that used in [**13**].

$$\frac{Du_g}{Dt} - fv_{ag} = 0, \frac{Dv_g}{Dt} + fu_{ag} = 0, \tag{3}$$

$$\frac{D\theta}{Dt} = 0, \tag{4}$$

$$\nabla.\mathbf{u} = 0,$$

$$\nabla p = \left(fv_g, -fu_g, -\frac{g\theta}{\theta_0}\right) \tag{5}$$

where

$$\begin{aligned} \mathbf{u} &\equiv (u,v,w) \equiv \mathbf{u_g} + \mathbf{u_{ag}}, \\ \mathbf{u_g} &\equiv (u_g, v_g, 0), \\ \frac{D}{Dt} &\equiv \frac{\partial}{\partial t} + \mathbf{u}.\nabla \end{aligned}$$

The notation is as in (1). Subscripts g and ag denote geostrophic and ageostrophic velocity (or wind) components respectively, where the geostrophic velocity is defined to be the horizontal component of velocity in balance with the pressure gradient. This definition is included in equation (5), as is the statement of hydrostatic balance, following (2).

We first solve the equations (for the velocity $\mathbf{u}$, potential temperature θ and pressure p) in the special case of f constant. We solve them in an open, bounded, connected set $\Omega \subset I\!R^3$ which has smooth boundary, with normal velocity $\mathbf{u.n}$ given on $\partial\Omega$. (In the meteorological literature the equations are sometimes posed on $\overline{\Omega}$.) For $\mathbf{x} = (x,y,z) \in \Omega$, by making the substitution

$$\mathbf{X} \equiv (X,Y,Z) \equiv (x + v_g/f, y - u_g/f, -(g/f^2\theta_0)\theta) \tag{6}$$

it is shown in Purser and Cullen [**18**] that we may replace (3) and (4) by

$$\frac{D\mathbf{X}}{Dt} = \mathbf{u_g}.$$

It is immediate from (5) and (6) that

$$\mathbf{X} = \nabla\left(\frac{p}{f^2} + \frac{1}{2}(x^2+y^2)\right). \tag{7}$$

(6) is known as the *geostrophic transformation* and is due to Hoskins [**13**].

We seek the values $\mathbf{X}$ which represent the state of the fluid; from these we can recover the physical quantities $\mathbf{u}$, θ and p. We think of $\mathbf{X}$ as a function of the physical space co-ordinates $\mathbf{x}$. Rewriting in terms of $\mathbf{X}$ and $\mathbf{x}$, we have

$$\frac{D\mathbf{X}}{Dt} = f(J.(\mathbf{X} - \mathbf{x})) \tag{8}$$

where

$$J = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We introduce the Lagrangian coordinates

$$\tilde{\mathbf{X}}(t, \mathbf{x}) = \mathbf{X}(t, \chi(t, \mathbf{x}))$$

where

$$t \to \chi(t, \mathbf{x})$$

is the trajectory of the fluid particle which is at $\mathbf{x}$ initially. Writing $P = p/f^2 + (x^2 + y^2)/2$, the Lagrangian form of (7) and (8) is

$$\tilde{\mathbf{X}}(t, \mathbf{x}) = \nabla P(t, \chi(t, \mathbf{x})), \tag{9}$$

$$\frac{\partial \tilde{\mathbf{X}}}{\partial t} = f(J.(\tilde{\mathbf{X}} - \chi)). \tag{10}$$

3 Solutions of the equations as a sequence of minimum energy states

In this section, for constant Coriolis parameter f, we describe a procedure which is equivalent to solving the semigeostrophic equations. At each fixed time t, $\mathbf{X}$ is predicted on particles (i.e. $\tilde{\mathbf{X}}(t, .)$ is predicted) using the evolution equation for $\mathbf{X}$. Then the particles are arranged according to an energy minimisation which is equivalent to a Monge mass transport problem. We show that this process is well posed, and yields that $\mathbf{X}(t, .) = \nabla\Psi(t, .)$ for some convex function $\Psi(t, .)$. Identifying Ψ with $p/f^2 + (x^2 + y^2)/2$, we can interpret solutions of the semigeostrophic equations as a sequence of minimum energy states. In the special case when $\tilde{\mathbf{X}}(t, .)$ is nondegenerate (in a sense made precise in section 3.3) for each t, we can find the trajectory mapping χ.

3.1 The Cullen-Norbury-Purser minimum energy principle

Solutions of the semigeostrophic equations (in the $\mathbf{X}$ variables) satisfy the evolution equation (8) subject to the constraint that the balance conditions (5) hold. Cullen, Norbury and Purser [7] interpreted solutions which are stable to small displacements of particles as follows: (8) says $\mathbf{X}$ evolves following fluid particles with tendency $f(y - Y, X - x, 0)$ and the fluid particles are

then rearranged conserving $\mathbf{X}$ so that (5) is satisfied and $p/f^2 + (x^2+y^2)/2$ is convex. (We shall see later that the latter condition corresponds to stability of the flow with respect to small displacements of particles.) Following Cullen, Norbury and Purser [7], we show that the above is equivalent to the following procedure. At each fixed time t, predict $\mathbf{X}$ on particles using (8). (In the Lagrangian variables defined above, we predict $\tilde{\mathbf{X}}(t,.)$ but not $\chi(t,.)$.) Now apply the *Cullen-Norbury-Purser* principle which states that for a solution, the particles are arranged to minimise geostrophic energy. The geostrophic energy E is defined as

$$\begin{aligned} E &= \int_\Omega \frac{1}{2}u_g^2 + \frac{1}{2}v_g^2 + \frac{g\theta z}{\theta_0} d\lambda_3(\mathbf{x}) \\ &= f^2\frac{1}{2}\int_\Omega X^2 + x^2 + Y^2 + y^2 d\lambda_3(\mathbf{x}) - f^2\int_\Omega \mathbf{x}.\mathbf{X} d\lambda_3(\mathbf{x}), \end{aligned}$$

where λ_3 denotes 3-dimensional Lebesgue measure. Now the change in energy under a volume preserving displacement which conserves $\mathbf{X}$ on particles satisfies

$$\begin{aligned} \delta E &= -\frac{f^2}{2}\int_\Omega \delta\mathbf{x}.\mathbf{X} d\lambda_3(x), \qquad (11) \\ \nabla.\delta\mathbf{x} &= 0. \end{aligned}$$

If $E = E(\mathbf{X})$ is at a minimum with respect to volume preserving displacements which conserve $\mathbf{X}$ on particles, then $\delta E = 0$, and the above equations imply that $\mathbf{X} = \nabla P$, for some scalar field P. Following Shutts and Cullen [19], write $\delta E = -f^2/2\int_\Omega \delta\mathbf{x}.(\mathbf{X} - \nabla P) d\lambda_3(x)$. Then, using $\delta(\nabla P) = (\delta\mathbf{x}.\nabla)\nabla P$, we have

$$\delta^2 E = \frac{f^2}{2}\int_\Omega ((\delta\mathbf{x}.\nabla)\nabla P).\delta\mathbf{x} d\lambda_3(x) \qquad (12)$$

so that the condition for a minimum is that P is a convex function. In [19] it is shown that this corresponds to the stability condition for the flow against small displacements $\delta\mathbf{x}$ in a frozen pressure field. Identifying P with $p/f^2 + (x^2+y^2)/2$, we have shown that the Cullen-Norbury-Purser principle implies that (5) holds, and that $p/f^2 + (x^2+y^2)/2$ is convex.

It remains to show that the Cullen-Norbury-Purser principle is well posed i.e. demonstrate that an energy minimiser exists and is unique. To translate the Cullen-Norbury-Purser principle into a variational problem, we use the concept of rearrangement of functions. An intuitive idea of when a function is a rearrangement of another function is as follows. Let Ω be as above, and let $f_1 : \Omega \to \mathbb{R}^3$ be a function. Imagine that Ω is a continuum of infinitesimal particles. Suppose that we exchange the particle positions, with each particle retaining its value of f_1. This yields a new function f_2, which is a *rearrangement* of f_1. Formally

Definition Let $f_1, f_2 : \Omega \to \mathbb{R}^3$, where Ω is as above, be integrable functions. Then f_1 and f_2 are *rearrangements* of each other if

$$\lambda_3\left(f_1^{-1}(B)\right) = \lambda_3\left(f_2^{-1}(B)\right)$$

for every Borel set $B \subset \mathbb{R}^3$. Rearrangement is an equivalence relation on the space of integrable functions, therefore we can define the set of rearrangements (or equivalence class) of a given vector valued function. For a prescribed f_0 we denote this set $\mathcal{R}(f_0)$. For any subdivision of Ω (into equal volumes) the set of rearrangements of f_0 includes every function that can be written as the composition of f_0 with a permutation of the subdivision. However it includes the composition of f_0 with any measure-preserving mapping, including those that are not injective. Moreover for $f_1, f_2 \in \mathcal{R}(f_0)$, there need not exist a measure-preserving mapping s such that $f_1 = f_2 \circ s$.

Returning to the Cullen-Norbury-Purser principle, suppose one possible state of the fluid at a fixed time t is described by a function $\mathbf{X_0} = (X_0, Y_0, Z_0)$. (One way to calculate a candidate state of the fluid would be to solve equation (8) on particles, i.e. solve (10) for $\tilde{\mathbf{X}}(t, .)$, and then assume the particles at time t are in the same configuration that they are in at time 0.) Values are fixed on particles, but the particles may take up any configuration, therefore possible states of the fluid are described by elements of $\mathcal{R}(\mathbf{X_0})$. The Cullen-Norbury-Purser principle states that we should minimise the geostrophic energy over all possible states, therefore we study the energy minimisation problem

$$\inf_{\mathbf{X} \in \mathcal{R}(\mathbf{X_0})} E(\mathbf{X}), \tag{13}$$

where the energy minimiser (if it exists and is unique) gives the actual state of the fluid. We make the additional (physically reasonable) assumption that $\mathbf{X_0} \in L^p(\Omega, \lambda_3, \mathbb{R}^3)$, where $2 \leq p < \infty$. This ensures that (13) is finite. In the next subsection we prove that the infimum in (13) is uniquely attained by a rearrangement equal to the gradient of a convex function. At each time t, we identify the convex function with $p/f^2 + (x^2 + y^2)/2$. In this way, solutions of the semigeostrophic equations can be viewed as a sequence of minimum energy states.

We note that in simple examples equation (8) may be easy to solve analytically, and in such cases the methods of Cullen, Norbury and Purser described above have been used to calculate solutions to the semigeostrophic equations. In general, however, this is not the case: we discuss solution of a closed form of the semigeostrophic equations in section 4.

Benamou [**1**] interprets the minimum energy principle in the following way: assume that the configuration of the fluid at each time t is an energy minimum with respect to permutation of the fluid particles. Recalling from (7) that $\mathbf{X} = \nabla P$, where $P = p/f^2 + (x^2 + y^2)/2$, and using the methods of Brenier [**3**, Proposition 2.1], it can be shown that P must be convex. The

approach used here differs in that we prove that an energy minimiser exists (and is unique), rather than assuming it exists.

3.2 Existence and uniqueness of an energy minimiser

In this subsection we show that the infimum in the variational problem (13) is uniquely attained by a rearrangement equal to the gradient of a convex function. For notational convenience, in this subsection we set the constant f^2 equal to 1. We make the following preliminary observation.

Lemma 1 *Let $\mathbf{X_0} \in L^p(\Omega, \lambda_3, \mathbb{R}^3)$ for $2 \leq p < \infty$, where $\Omega \subset \mathbb{R}^3$ is an open, bounded and connected set with smooth boundary, and λ_3 denotes 3-dimensional Lebesgue measure. Let $\mathbf{X_1} \in \mathcal{R}(\mathbf{X_0})$. Then*

$$\int_\Omega x^2 + X_1^2 + y^2 + Y_1^2 d\lambda_3(\mathbf{x}) = \int_\Omega x^2 + X_0^2 + y^2 + Y_0^2 d\lambda_3(\mathbf{x}),$$

and

$$\int_\Omega z^2 + Z_1^2 d\lambda_3(\mathbf{x}) = \int_\Omega z^2 + Z_0^2 d\lambda_3(\mathbf{x})$$

where $\mathbf{X_0} = (X_0, Y_0, Z_0)$ and $\mathbf{X_1} = (X_1, Y_1, Z_1)$.

Proof. $\mathbf{X_1} \in \mathcal{R}(\mathbf{X_0})$ therefore it follows (from for example Douglas [**9**, Theorem 2.2]) that

$$\int_\Omega X_1^2 d\lambda_3(\mathbf{x}) = \int_\Omega X_0^2 d\lambda_3(\mathbf{x}).$$

A similar result holds for Y_0 and Y_1, Z_0 and Z_1. The result follows.

¿From Lemma 1 it is easily seen that any $\hat{\mathbf{X}} \in \mathcal{R}(\mathbf{X_0})$ attains the infimum in (13) if and only if $\hat{\mathbf{X}}$ is a maximiser of $\int_\Omega \mathbf{x}.\mathbf{X}(\mathbf{x}) d\lambda_3(\mathbf{x})$ over $\mathbf{X} \in \mathcal{R}(\mathbf{X_0})$, and a minimiser of $\int_\Omega |\mathbf{X}(\mathbf{x}) - \mathbf{x}|^2 d\lambda_3(\mathbf{x})$ over $\mathbf{X} \in \mathcal{R}(\mathbf{X_0})$, where $|.|$ denotes Euclidean distance on $\mathbb{R}^3$. We study the latter problem first.

The minimisation problem

$$\inf_{\mathbf{X} \in \mathcal{R}(\mathbf{X_0})} \int_\Omega |\mathbf{X}(\mathbf{x}) - \mathbf{x}|^2 d\lambda_3(\mathbf{x}) \tag{14}$$

may be rewritten as a Monge mass transport problem as follows. Define, for Borel sets $B \subset \mathbb{R}^3$, $\nu_1(B) = \lambda_3(B \cap \Omega)/\lambda_3(\Omega)$, and $\nu_2(B) = \lambda_3(\mathbf{X_0}^{-1}(B))/\lambda_3(\Omega)$. Then ν_1 and ν_2 are Borel probability measures on $\mathbb{R}^3$: let S denote the set of measure-preserving mappings from $(\mathbb{R}^3, \nu_1)$ to $(\mathbb{R}^3, \nu_2)$. (To link with terminology used elsewhere in this volume, $s \in S$ if and only if s pushes ν_1 forward to ν_2, i.e. $s_\# \nu_1 = \nu_2$.) Ignoring the ν_1 negligible set $\mathbb{R}^3 \setminus \Omega$, we have $\mathbf{X} \in \mathcal{R}(\mathbf{X_0})$ if and only if $\mathbf{X} \in S$. It follows that (14) is equivalent to

$$\inf_{s \in S} \int_{\mathbb{R}^3} |s(\mathbf{x}) - \mathbf{x}|^2 d\nu_1(\mathbf{x}) \tag{15}$$

It has been shown that (15) has a unique minimiser by proving the existence of a unique minimiser (which has a particular form) of an appropriate Monge-Kantorovich problem: a lucid explanation can be found in Gangbo and McCann [**10**]. The minimising rearrangement is the *monotone rearrangement* of $\mathbf{X_0}$, denoted $\mathbf{X_0}^{\#}$: it is equal to the gradient of a convex function (defined on $\mathbb{R}^3$) ν_1 almost everywhere, and is the (essentially) unique rearrangement of $\mathbf{X_0}$ which may be characterised in this way. (The existence and uniqueness of the monotone rearrangement is due to Brenier [**3**]: if we relax the assumptions on Ω to having finite Lebesgue measure, then the fact that the monotone rearrangement exists and is unique is due to McCann [**16**].) It is important to note that we make no restriction on ν_2. If ν_2 is assumed to be absolutely continuous with respect to Lebesgue measure, this is equivalent to $\mathbf{X_0}$ satisfying the *nondegeneracy* condition of Brenier [**3**, Theorem 1.2], that is $\lambda_3(\mathbf{X_0}^{-1}(E)) = 0$ for every set $E \subset \mathbb{R}^3$ with negligible Lebesgue measure. Nondegenerate functions have no level sets of positive measure, so such an assumption would exclude some physically reasonable solutions.

Alternatively we seek maximisers of $\int_\Omega \mathbf{x}.\mathbf{X}(\mathbf{x})d\lambda_3(\mathbf{x})$ over $\mathbf{X} \in \mathcal{R}(\mathbf{X_0})$. We consider the Dual Monge-Kantorovich problem of Brenier [**3**, p. 396]. (Douglas [**9**, Theorem 2.2] proved that the definition of rearrangement in Brenier [**1**] is equivalent to the one given above.) Proposition 3.1 (iii) and (v) and Theorem 3.1 of Brenier [**3**] yield that the problem has a unique solution with a particular form. It may be deduced that

$$\int_\Omega \mathbf{x}.\mathbf{X_0}^{\#}(\mathbf{x})d\lambda_3(\mathbf{x}) > \int_\Omega \mathbf{x}.\mathbf{X}(\mathbf{x})d\lambda_3(\mathbf{x})$$

for each $\mathbf{X} \in \mathcal{R}(\mathbf{X_0})\backslash\{\mathbf{X_0}^{\#}\}$. Douglas [**9**, Theorem 3.1] showed that this result could also be recovered from Brenier [**3**, Theorems 1.1, 1.2 and Proposition 2.1].

3.3 Identification of the trajectory mapping via polar factorisation

The previous two subsections have shown that for a (stable) solution to the semigeostrophic equations, the methods of Cullen, Norbury and Purser [**7**] yield that at each fixed time t, $\mathbf{X}(t,.) = \nabla\Psi(t,.)$ for some convex function $\Psi(t,.)$ which we identify with $p/f^2 + (x^2+y^2)/2$. In this subsection we study the Lagrangian version of the problem, that is equations (9) and (10), using the techniques of Benamou and Brenier [**1**, **2**]. Our aim is to find the trajectory mapping χ.

For each t, assume $\tilde{\mathbf{X}}(t,.)$ is square integrable and nondegenerate (in the sense of Brenier [**3**], that is $\lambda_3(\tilde{\mathbf{X}}^{-1}(t,E)) = 0$ for every set $E \subset \mathbb{R}^3$ of zero Lebesgue measure). Then, for each t, Brenier [**3**, Theorem 1.2] yields that $\tilde{\mathbf{X}}(t,.)$ has a unique polar factorisation into the composition of its monotone rearrangement with a measure-preserving mapping, that is there exists a

convex function $\Psi(t,.)$, and a unique measure-preserving mapping $s(t,.)$ such that

$$\tilde{\mathbf{X}}(t,.) = \nabla\Psi(t, s(t,.)). \tag{16}$$

We identify Ψ with $p/f^2 + (x^2+y^2)/2$ (as before), and s with the trajectory mapping χ. Thus we have found the trajectory mapping at each fixed time. For each t, Brenier [**3**, Theorem 1.2] yields that $\chi(t,.)$ is the projection of $\tilde{\mathbf{X}}(t,.)$ onto the set of measure-preserving mappings from Ω to Ω. Denoting this projection mapping by Π_S, (10) yields that

$$\frac{\partial\tilde{\mathbf{X}}}{\partial t} = f(J.(Id - \Pi_S)\tilde{\mathbf{X}}) \tag{17}$$

where Id denotes the identity map on Ω. Benamou [**1**] refers to (17) as the *dynamic rearrangement equation.*

However as we noted in the previous subsection, there is no physical reason why $\tilde{\mathbf{X}}(t,.)$ should be nondegenerate. Suppose that $\tilde{\mathbf{X}}(t,.)$ has at least one level set of positive measure, and that $\tilde{\mathbf{X}}(t,.)$ satisfies (16) for some convex function $\Psi(t,.)$ and some measure-preserving mapping $s(t,.)$. $\nabla\Psi(t,.)$ has a corresponding level set of positive measure, and we can choose a non-trivial measure–preserving transformation (i.e. an invertible measure-preserving mapping which maps measurable sets to measurable sets) τ : $\Omega \to \Omega$ which leaves points other than those in the level set fixed. Now $\tau \circ s(t,.)$ is a measure-preserving mapping which is different from $s(t,.)$, and satisfies $\tilde{\mathbf{X}}(t,.) = \nabla\Psi(t, \tau \circ s(t,.))$: hence in this case we cannot define the trajectory mapping uniquely without further information. Some extensions of Brenier's [**3**] results on existence and uniqueness of a polar factorisation of an integrable vector-valued function are proved in Burton and Douglas [4]. A question of current interest is how to choose the trajectory mapping when the polar factorisation (16) is known to exist but is not unique.

4 Solution of the time evolution problem

In this section we address the problem of proving existence of solution for the semigeostrophic equations. The previous section yielded that for each time t, $\mathbf{X}(t,.) = \nabla P(t,.)$ for some convex function $P(t,.)$. We assume that $\mathbf{X}(t,.)$ is nondegenerate for each t, and rewrite the semigeostrophic equations as a coupled Monge-Ampère /transport problem in section 4.1. We discuss results by Benamou and Brenier [**1**,**2**], and an alternative strategy for proving existence of solutions. The second half of this section considers properties of the long term evolution of the semigeostrophic equations: the important questions are illustrated by studying a simpler model.

4.1 The semigeostrophic equations formulated as a coupled Monge-Ampère/transport problem

In the previous section we showed that solutions of the semigeostrophic equations can be viewed as a sequence of minimum energy states: however we did not address the problem of showing that solutions exist. Our aim in this subsection is to rewrite the semigeostrophic equations as a coupled Monge-Ampère/ transport problem, and discuss existence of solution for this formulation. Previous work showed that for each t, $\mathbf{X}(t,.) = \nabla P(t,.)$ for some convex function $P(t,.)$. We assume that $\mathbf{X}(t,.)$ is nondegenerate for each t, and show that the (Legendre–Fenchel) conjugate convex function of $P(t,.)$ satisfies a Monge-Ampère problem in a generalised sense. (This replaces (7) and the above results concerning $\mathbf{X}(t,.)$.) We rewrite (8) as a transport equation, and discuss the resulting coupled problem, which is a closed form of the semigeostrophic equations. An area of current interest is to develop the corresponding theory for $\mathbf{X}(t,.)$ which fail the nondegeneracy condition.

The methods of section 3 yielded $\mathbf{X}$ as a function of $\mathbf{x}$ at each fixed time t; we now wish to find a mapping from $\mathbf{X}$ to $\mathbf{x}$. For each t, given $\mathbf{X}(t,.) = \nabla P(t,.)$, essentially we restrict attention to the case where $\nabla P(t,.)$ is injective, so that we have $\mathbf{x}(t,.) = (\nabla P(t,.))^{-1}$. For each t suppose $\mathbf{X}(t,.) : \Omega \to I\!R^3$ is an integrable function which satisfies $\mathbf{X}(t,.) = \nabla P(t,.)$ almost everywhere for some proper lower semicontinuous convex function $P(t,.) : I\!R^3 \to I\!R \cup \{\infty\}$, and further suppose that $\mathbf{X}(t,.)$ is nondegenerate in the sense of Brenier [3]. Burton and Douglas [4, Lemma 5] yields that $\mathbf{X}(t,.)$ is injective (except possibly on a set of zero measure). Moreover for each fixed t standard convex analysis demonstrates that $\mathbf{x}(t,.) = \nabla R(t,.)$, and

$$P(t,\mathbf{x}) + R(t,\mathbf{X}(t,\mathbf{x})) = \mathbf{x}.\mathbf{X}(\mathbf{x},t) \tag{18}$$

for almost every $\mathbf{x} \in \Omega$, where $R(t,.)$ denotes the (Legendre–Fenchel) *conjugate convex* function of $P(t,.)$, that is

$$R(t,\mathbf{X}) = \sup\{\mathbf{x}.\mathbf{X} - P(t,\mathbf{x}) : \mathbf{x} \in I\!R^3\}.$$

(We note that (18) still holds for $\mathbf{X}(t,.)$ which fail the nondegeneracy condition.)

For each t let $\mathbf{X}(t,.)$ be as above, and define a Borel measure ν_t by $\nu_t(B) = \lambda_3(\mathbf{X}^{-1}(t,B))$ for Borel subsets B of $I\!R^3$. $\mathbf{X}(t,.)$ is nondegenerate by assumption, therefore $\nu_t(E)$ is well defined for Lebesgue subsets E of $I\!R^3$, and moreover ν_t is absolutely continuous with respect to Lebesgue measure; let $\rho(t,.)$ denote the Radon–Nikodym derivative of ν_t with respect to Lebesgue measure. Then

$$\int_{I\!R^3} \rho(t,\mathbf{X}) d\lambda_3(\mathbf{X}) = \int_{I\!R^3} d\nu_t = \lambda_3(\Omega),$$

and Brenier [3] shows that $R(t,.)$ satisfies the following Monge-Ampère problem (in a generalised sense made precise in [3]) for each t:

$$\det H(R(t,\mathbf{X})) = \rho(t,\mathbf{X}), \tag{19}$$
$$\nabla R(t,\mathbf{X}) \text{ maps the support of } \rho(t,\mathbf{X}) \text{ into } \Omega,$$

where $R(t,.)$ is as above, and H denotes the Hessian matrix. The evolution problem (8) can be rewritten as

$$\frac{\partial \rho}{\partial t} + \nabla.(\rho \mathbf{U}) = 0 \tag{20}$$

where $\mathbf{U} = f(y-Y, X-x, 0)$ and $\nabla \equiv (\partial/\partial X, \partial/\partial Y, \partial/\partial Z)$. Using (8), and recalling that $\mathbf{x}(t,.) = \nabla R(t,.)$ for each t, we have

$$\mathbf{U} = f(J.(Id - \nabla R)) \tag{21}$$

where Id denotes the identity mapping on $\mathbb{R}^3$. The transport equation (20) is coupled to the Monge-Ampère problem (19) by (21), where $R(t,.)$ is as above. Manipulations based on (18) show that if $\mathbf{U}$ is sufficiently smooth, $\nabla.\mathbf{U} = 0$. When this holds, the solution of (20) can be viewed as rearranging the initial values of ρ i.e. $\rho(t,.)$ is a rearrangement of $\rho(0,.)$ for each positive time t.

Benamou [**1**, Chapter 2] showed the existence of solutions on a finite time interval for (19) and a weak formulation of (20), coupled as described above. His approach was to show existence of solutions for approximate problems and then pass to the limit. Benamou and Brenier [**2**] proved a result of global weak existence. The authors are not aware of any uniqueness results. An alternative strategy to prove existence (and possibly uniqueness) of solutions for the coupled problem is to use a fixed point argument. If we could show that (19) yields that $R(t,.) \in W^{2,1}$, then (21) gives that $\mathbf{U}(t,.) \in W^{1,1}$, and we could use the transport theory of DiPerna and Lions [**8**]DiPerna and Lions [**8**] to solve (20). However the authors are not aware of any proof of this result. The transport equation can still be solved if the velocity is (slightly) less regular. A suitable condition on $\mathbf{U}$ to allow us to solve (20) is that

$$\int_{|\mathbf{X}|\leq K} |\nabla \mathbf{U}(t,\mathbf{X})| d\mathbf{X} \leq C_K(t), \forall K < \infty \text{ with } C_K \in L^1_{\text{loc}}(\mathbb{R}^+) \tag{22}$$

This corresponds to the limiting case of the transport theory of DiPerna and Lions [**8**]. The challenge is to use (19) (which is satisfied in a generalised sense) to prove that R is sufficiently regular to enforce this. If we assume that Ω and the support of $\rho(t,.)$ are convex, Caffarelli [**5**] shows that if $\rho(t,.) \in C^\alpha$, then $R(t,.) \in C^{2,\alpha}$: using (21) it follows that $|\nabla \mathbf{U}(t,.)| \in C^\alpha$. (However even if the support of $\rho(0,.)$ is convex, there is no reason why the support of $\rho(t,.)$ should be convex for each positive time t.) More recent results by Caffarelli elsewhere in this volume suggest that the condition (22) can be met.

4.2 Long time evolution properties

We first illustrate the key questions by using the 2-dimensional barotropic vorticity equations.

$$\begin{aligned} \frac{D}{Dt}(\zeta + f) &= 0 \\ \frac{D}{Dt} \equiv \frac{\partial}{\partial t} - \frac{\partial \psi}{\partial y}\frac{\partial(\zeta + f)}{\partial x} + \frac{\partial \psi}{\partial x}\frac{\partial(\zeta + f)}{\partial y} & \\ \nabla^2 \psi &= \zeta \end{aligned} \tag{23}$$

Here, ζ is the vorticity, and ψ the streamfunction. Proving regularity of the solutions of this equation requires establishing bounds on the velocity gradients. Such bounds are used to show that the vorticity is bounded by its initial values in the time evolution, and hence existence of regular solutions to the equations. However, if a maximum norm is used, the bound on the velocity gradients depends on the vorticity gradients (Gerard [11],p 425):

$$||\nabla \mathbf{U}(t)|| \leq C \log(2 + ||\nabla(\zeta + f)(t)||) \tag{24}$$

Exact definitions of the norms used are given in Gerard, but they are essentially maximum norms. The bound is derived from the solution procedure for the Poisson equation in (23) for ψ in terms of ζ. The vorticity gradients can be estimated by using (23) to give the following equation for the time evolution of the vorticity gradients:

$$\frac{D}{Dt}\nabla(\zeta + f) + \begin{pmatrix} -\frac{\partial^2 \psi}{\partial x \partial y} & \frac{\partial^2 \psi}{\partial x^2} \\ -\frac{\partial^2 \psi}{\partial y^2} & \frac{\partial^2 \psi}{\partial y \partial x} \end{pmatrix} \nabla(\zeta + f) = 0. \tag{25}$$

These equations can be used along with (24) to estimate the rate of increase of vorticity gradients. This estimate allows exponential growth in time. This does not prevent regularity being proved for all time, but allows the accumulation of enstrophy (variance of vorticity) at small scales. This can be expressed as a statement that

$$||(\nabla \mathbf{U})^2|| \leq C||\mathbf{U}^2|| \tag{26}$$

where C grows exponentially in time. If the L^2 norm is used instead of the maximum norm, then an estimate of the form (26) holds with C independent of time, but dependent on the domain size. Thus the mean scale of the flow is bounded, but local regions where small scales are generated are permitted. This agrees with widespread computational experience.

Now consider the semigeostrophic model, as expressed by (20) and (19). An equation of the form (25) still governs the rate of growth of gradients of

ρ. Defining $\Psi = \frac{1}{2}(X^2 + Y^2) - R$, we obtain

$$\frac{D}{Dt}(\nabla\rho) + \begin{pmatrix} -\frac{\partial^2\Psi}{\partial X\partial Y} & \frac{\partial^2\Psi}{\partial X^2} \\ -\frac{\partial^2\Psi}{\partial Y^2} & \frac{\partial^2\Psi}{\partial Y\partial X} \end{pmatrix} \nabla\rho = 0 \tag{27}$$

The equation that determines Ψ from ρ is the Monge-Ampère equation (19) for R, followed by addition of the fixed function $\frac{1}{2}(X^2 + Y^2)$ to obtain Ψ. Thus, in effect, ρ is transported by a convex streamfunction. If the flow is steady, convexity prevents exponential growth of gradients of ρ, except for time periods short compared with that needed for a particle to make a complete circuit of a streamline. Only algebraic growth is permitted. For general time dependent solutions of (20) and (19), it is a plausible but unproved conjecture that exponential growth is prevented and a sharper estimate than (24) obtained. If so, there would be no systematic enstrophy cascade in semigeostrophic turbulence. This would provide an explanation of why weather systems are observed to look qualitatively the same all the time, with no systematic tendency for migration of enstrophy to the smallest scales.

5 Semigeostrophic theory with variable f

Thus far we have studied the semigeostrophic equations in the special case of constant Coriolis parameter f. This restricts the equations to regions where this assumption is reasonable, i.e. regions where the variation in latitude is small. However the equations are asymptotically valid where the horizontal length scale is comparable to the radius of the earth. In this regime the correct variation of f must be included. Sections 3 and 4 have demonstrated the utility of the geostrophic transformation (6) for studying the semigeostrophic equations on an f plane: in this section we seek a corresponding transformation for semigeostrophic theory where f varies. In section 5.1 we show that the geostrophic transformation (6) can be derived via energy minimisation: motivated by this result, we define a transformation for semigeostrophic theory with variable f by minimising an appropriate energy functional. The remainder of section 5 is concerned with formulating this problem and showing that a unique minimiser exists.

In section 3 we showed that solutions of the semigeostrophic equations on an f plane could be interpreted in the following way: at each fixed time t, predict $\mathbf{X}$ on particles using evolution equation (8) (or more precisely predict $\tilde{\mathbf{X}}(t,.)$ using (10)) and then minimise geostrophic energy over the set of rearrangements of one possible state of the fluid. For semigeostrophic theory with variable f we address the analogous energy minimisation problem in this section, but not the formulation of the evolution problem. This section contains material that will appear in Cullen, Douglas and Roulstone [**6**]:

the question of how to combine the transformation with the semigeostrophic equations will be addressed in [6].

5.1 Defining a geostrophic transformation via energy minimisation

We begin by showing that the geostrophic transformation (6) for the semigeostrophic equations on an f plane can be obtained via the energy minimisation problem (13). Suppose we have a mapping from particle label space to $\mathbf{X} = (X, Y, Z)$ variables which is defined by requiring that volume preserving displacements which conserve $\mathbf{X}$ on particles are exactly those displacements which conserve potential temperature and geostrophic absolute momentum on particles. It follows that for a given $\mathbf{x} = (x, y, z) \in \Omega$, $f^2(y-Y)^2/2 + f^2(X-x)^2/2 - f^2zZ$ is equal to $u_g^2/2 + v_g^2/2 + (g\theta z)/\theta_0$. $\mathbf{X}$ is known on particles: let one possible configuration of the particles yield a function $\mathbf{X_0}$. Further suppose that $\mathbf{X}$ is the unique minimiser of

$$\inf_{\mathbf{X}\in\mathcal{R}(\mathbf{X_0})} E(\mathbf{X}) = \inf_{\mathbf{X}\in\mathcal{R}(\mathbf{X_0})} f^2 \int_\Omega \frac{1}{2}\left\{(y-Y)^2 + (X-x)^2\right\} - zZd\lambda_3(\mathbf{x}).$$

The results of section 3 yield that $\mathbf{X}$ is well defined, and further $\mathbf{X} = \nabla\Psi$ for some convex function Ψ. Moreover $E(\mathbf{X})$ is the geostrophic energy of the flow, and it is at a minimum with respect to volume preserving displacements which conserve $\mathbf{X}$ on particles. Identifying Ψ with $p/f^2 + (x^2+y^2)/2$, this implies that the geostrophic and hydrostatic relations (5) are satisfied, therefore we obtain

$$\begin{aligned} \mathbf{X} = \nabla\Psi &= \nabla\left(\frac{p}{f^2} + \frac{x^2+y^2}{2}\right) \\ &= \left(x + \frac{v_g}{f}, y - \frac{u_g}{f}, -\frac{g\theta}{f^2\theta_0}\right), \end{aligned}$$

the geostrophic transformation (6).

We now return to the basic equations (1) and allow the Coriolis parameter f to vary in space. Define the energy integral

$$E = \int_D \frac{1}{2}(u^2+v^2) + \frac{g\theta z}{\theta_0},$$

where D is the domain of the fluid, and consider the conditions for E to be stationary with respect to volume preserving particle displacements $\delta\mathbf{x}$ satisfying

$$\begin{aligned} \delta u &= f\delta y, \\ \delta v &= -f\delta x, \\ \delta\theta &= 0. \end{aligned} \tag{28}$$

Then

$$\delta E = \int_D f u \delta y - f v \delta x + \frac{g\theta \delta z}{\theta_0}. \tag{29}$$

Using the condition that the displacement is volume preserving, $\nabla.\delta\mathbf{x} = 0$, gives

$$(fv, -fu, -\frac{g\theta}{\theta_0}) = \nabla s \tag{30}$$

for some scalar s. (30) corresponds exactly to the geostrophic and hydrostatic relations (2) if we set $s = p$. (When the geostrophic relation holds, $u = u_g$ and $v = v_g$.) Thus we have identified stationary points of the energy with geostrophic and hydrostatic balance in the variable f case.

We seek a new definition of a coordinate $\mathbf{X}$ for semigeostrophic theory with variable f, and by analogy with the results presented above for the constant f case, $\mathbf{X}$ will be defined on particles so that displacements which conserve $\mathbf{X}$ on particles are equivalent to those satisfying (28) with $u = u_g$ and $v = v_g$: then $\mathbf{X}$ will be found by minimising an appropriate energy functional over of the set of rearrangements of a given function. We work on the sphere S^2. On the rotating Earth, $f = 2\Omega \sin\phi$, where ϕ is the latitude and Ω the angular velocity of the Earth's rotation.

Definition. A displacement under the action of the Coriolis effect is a volume preserving displacement $\delta\mathbf{x}$ which satisfies

$$\delta u_g = f\delta y, \delta v_g = -f\delta x, \delta\theta = 0. \tag{31}$$

The *potential position* of a particle satisfies the condition that if a particle is at rest at its potential position, and is then displaced under the action of the Coriolis effect to its real position, the particle will then acquire its observed velocity.

For a given $\mathbf{x}$, $\mathbf{X}(\mathbf{x})$ is chosen to be the potential position of $\mathbf{x}$. For simplicity, we ignore the potential energy term in the geostrophic energy when we formulate the energy minimisation problem. (We discuss the consequences of this in the concluding paragraph of this subsection.) We define a "distance function" d such that $d(\mathbf{x}, \mathbf{X})^2$ is equal to $u_g^2 + v_g^2$. Then

$$\frac{1}{2}\int d(\mathbf{x}, \mathbf{X})^2$$

is equal to the (geostrophic) kinetic energy of the flow. For the f-plane (when the equations are defined on a subset of $\mathbb{R}^3$) the distance is just the Euclidean distance (in $\mathbb{R}^2$) between the first two components of $\mathbf{x}$ and $\mathbf{X}$, multiplied by f. When f varies, guided by the above and (31), we would like to define a distance function by integrating f from $\mathbf{x}$ to $\mathbf{X}$ along an appropriate path in S^2. However relations (31) are non-holonomic. Instead we interpret (31) as a rescaling of the infinitesimal distances. In the next

subsection we introduce a conformal rescaling of the sphere: on this manifold the distance induced by the metric is the "distance function" we seek. To find the relationship between $\mathbf{x}$ and $\mathbf{X}$, by analogy to the f-plane case discussed above we minimise energy over all possible flows, i.e. consider

$$\inf_{\mathbf{X}\in\mathcal{R}(\mathbf{X}_0)} \frac{1}{2}\int d(\mathbf{x},\mathbf{X})^2$$

where $\mathbf{X}_0$ is a prescribed mapping (obtained from a possible particle configuration, noting that $\mathbf{X}$ is known on particles).

In the next section we give a precise formulation of the above problem, and then rewrite it as a Monge mass transport problem on a manifold. The simplification of ignoring the geostrophic potential energy restricts our results to semigeostrophic theory where the potential energy is conserved by $\mathbf{X} \in \mathcal{R}(\mathbf{X}_0)$. This is true for incompressible shallow water semigeostrophic theory on a sphere.

5.2 Formulation of the energy minimisation as a Monge problem on a manifold

In this section we calculate a coordinate transformation, which maps each point to its potential position, by finding the unique minimiser of an energy over all possible flows. Instead of working on the sphere S^2, we translate the problem to a conformal rescaling of the sphere. On this manifold the distance induced by the metric is exactly the distance function introduced in the previous section. In this setting we can apply a recent theorem concerning Monge mass transport problems on manifolds. For the conformal rescaling to make sense, we must treat the two hemispheres separately. To satisfy the hypotheses of the theorem, we choose an energy functional which represents the kinetic energy of the flow on a subset of the hemisphere. The coordinate transformation we obtain is only defined on this subset; we discuss the ramifications of this in the concluding paragraph.

To state the energy minimisation problem precisely, we need some definitions and notation. Let g_{ij} denote the metric for the Riemannian manifold S^2. Define $\hat{g}_{ij} = f^2 g_{ij}$, a conformal rescaling of the metric for $f > 0$. (Conformal mappings and related concepts are discussed in Jost [14].) Let S^+ denote the upper hemisphere excluding the equator, and write $\hat{S}^+$ for the conformal rescaling of S^+ under the Coriolis transformation. $(\hat{S}^+, \hat{g}_{ij})$ is a Riemannian manifold. Let d be the Riemannian distance induced on $\hat{S}^+$ by $\hat{g}_{ij}$. Let $\hat{\mu}$ denote surface area on $\hat{N} \subset \hat{S}^+$, where the closed set $\hat{N}$ will be chosen later. We extend our definition of rearrangement to mappings which take values on manifolds in the obvious way.

Definition. Let $f_1, f_2 : \hat{S}^+ \to \hat{S}^+$ be mappings such that $f_1^{-1}(B)$ and $f_2^{-1}(B)$ are $\hat{\mu}$-measurable for each Borel subset B of $\hat{S}^+$. Then f_1 is a

rearrangement of f_2 if

$$\hat{\mu}\left(f_1^{-1}(B)\right) = \hat{\mu}\left(f_2^{-1}(B)\right)$$

for every Borel subset B of $\hat{S}^+$.

We define our new coordinate $\hat{\mathbf{X}}$ to be the unique minimiser of

$$\inf_{\mathbf{X}\in\mathcal{R}(\mathbf{X_0})} \frac{1}{2}\int_{\hat{S}^+} d(\mathbf{x},\mathbf{X}(\mathbf{x}))^2 d\hat{\mu}(\mathbf{x}), \tag{32}$$

where $\mathbf{X_0} : \hat{S}^+ \to \hat{N}$ is a prescribed mapping such that the inverse image of a Borel subset of $\hat{S}^+$ is a Borel set. (32) represents the kinetic energy of the flow restricted to $\hat{N}$. (Mutatis mutandis we define the new co-ordinate on the conformally rescaled lower hemisphere in the same way.)

To demonstrate that (32) is uniquely attained, we rewrite (32) as a Monge mass transport problem, and then appeal to a recent theorem by McCann. Define the Borel measure $\hat{\nu}$ on $\hat{S}^+$ by $\hat{\nu}(B) = \hat{\mu}(\mathbf{X_0}^{-1}(B))$ for each Borel subset B of $\hat{S}^+$. Then S, the set of measure-preserving mappings from $(\hat{S}^+, \hat{\mu})$ to $(\hat{S}^+, \hat{\nu})$, is exactly $\mathcal{R}(\mathbf{X_0})$. We may rewrite (32) as

$$\inf_{s\in S} \frac{1}{2}\int_{\hat{S}^+} d(\mathbf{x}, s(\mathbf{x}))^2 d\hat{\mu}(\mathbf{x}). \tag{33}$$

The following theorem is due to McCann [**17**]. (Related results are discussed elsewhere in this volume.)

Theorem 2 *Let (M, g) be a smooth (C^2) Riemannian manifold of dimension 2, where*

$$\textit{length}\ (\sigma) = \int_0^1 \left(g_{ij}(\sigma(s)) \frac{d\sigma^i}{ds} . \frac{d\sigma^j}{ds} \right)^{1/2} ds$$

for a smooth curve $\sigma : [0,1] \to M$, and

$$d(x,y) = \inf\{\textit{length}\ (\sigma) : \sigma \textit{ a piecewise smooth curve connecting } x \textit{ and } y\}.$$

Let μ, ν be Borel probability measures on M, and suppose μ vanishes on all sets of zero area. Further suppose that each $\mathbf{x} \in$ spt μ can be connected to each $\mathbf{y} \in$ spt ν by a minimal geodesic. Let S denote the set of measure-preserving mappings from (M, μ) to (M, ν). Then if

$$\inf_{s\in S} \frac{1}{2}\int_M d(s(\mathbf{x}), \mathbf{x})^2 d\mu(\mathbf{x})$$

is finite, the infimum is uniquely (up to μ-negligible sets) attained by

$$s(\mathbf{x}) = \exp_{\mathbf{x}}(-\nabla\varphi(\mathbf{x}))$$

where φ is a $d^2/2$-concave function, and exp *denotes the exponential map on the manifold M.*

We define some terms used in the statement of the above theorem. The *support* of a Borel probability measure μ on M, denoted *spt* μ, refers to the smallest closed set $U \subset M$ which has full measure. The exponential mapping is a mapping from the tangent space of M at $\mathbf{x}$ to the manifold. There is a unique geodesic which starts at x and proceeds in the direction $-\nabla\varphi(\mathbf{x})$; $\exp_{\mathbf{x}}(-\nabla\varphi(\mathbf{x}))$ is the position reached when we have travelled distance $|-\nabla\varphi(\mathbf{x})|$ along this geodesic. A function $\varphi : M \to \mathbb{R} \cup \{-\infty\}$, not identically equal to $-\infty$, is said to be $d^2/2$-*concave* if there exists $A \subset M \times \mathbb{R}$ such that

$$\varphi(\mathbf{x}) = \inf_{(\mathbf{y},\alpha)\in A} \frac{d(\mathbf{x},\mathbf{y})^2}{2} + \alpha.$$

(The concept of $d^2/2$-concavity is an adaptation of the notion of concavity to the relevant geometry.)

In the special case where $M = \mathbb{R}^2$ and d is Euclidean distance, geodesics are straight lines. Let μ be planar Lebesgue measure restricted to a bounded measurable set Ω. The minimiser, which is defined on Ω, has the form $s(\mathbf{x}) = \mathbf{x} - \nabla\varphi(\mathbf{x}) = \nabla(|\mathbf{x}|^2/2 - \varphi(\mathbf{x}))$, and $d^2/2$-concavity of φ is equivalent to convexity of $|\mathbf{x}|^2/2 - \varphi(\mathbf{x})$ on $\mathbb{R}^2$. Thus in the case of the f-plane we recover the result that the minimiser is the gradient of a convex function. (Note however that we are only minimising the kinetic energy over the set of rearrangements of $\mathbf{X_0}$, that is studying the problem

$$\inf_{\mathbf{X}\in\mathcal{R}(\mathbf{X_0})} \frac{f^2}{2} \int_{\mathbb{R}^2} (X - x)^2 + (Y - y)^2 d\mu(\mathbf{x}).)$$

To satisfy the hypotheses of Theorem 2 we choose $\hat{N}$ to be a closed set having the property that each $\mathbf{x} \in \hat{N}$ is connected to every $\mathbf{y} \in \hat{N}$ by a minimal geodesic (lying in $\hat{S}^+$). Noting that we may renormalise $\hat{\mu}(\hat{S}^+)$ and $\hat{\nu}(\hat{S}^+)$ to 1, and that (33) is finite, Theorem 2 yields that (33), and hence (32), has a unique minimiser. However we note that the new coordinate transformation is only defined on $\hat{N}$. The manifold $(\hat{S}^+, \hat{g})$ is not geodesically complete, so we cannot choose $\hat{N}$ equal to $\hat{S}^+$. Areas of current interest are whether we can find a maximal $\hat{N}$ with the required properties, and how to define a coordinate transformation for $\hat{S}^+ \setminus \hat{N}$. The fact that we are not able to define a transformation for the whole of $\hat{S}^+$ is a feature we would anticipate from numerical calculations and observations of real weather. Computations by Mawson [**15**] show that there is a discontinuous change in the behaviour of solutions (of a semigeostrophic model) in a band around the equator. Moreover weather observations show that tropical cyclones do not form within five degrees of latitude of the equator.

Acknowledgments

The authors would like to thank the organisers of the NSF–CBMS Conference *The Monge-Ampère Equation: Applications to Geometry and Opti-*

misation for their hospitality and financial support. Furthermore they are grateful for the hospitality of the Isaac Newton Institute for Mathematical Sciences during the workshop *Atmosphere and Ocean Dynamics*, where they had stimulating discussions concerning this work. They would like to thank the referee for drawing their attention to the work of Benamou and Brenier [1,2], and furthermore they are grateful to R.J. McCann for his insightful comments. The second author's research was also supported by EPSRC Research Grant reference GL/L43220.

6 References

1. J.-D. BENAMOU, *Transformation conservant la measure, mécanique des fluides incompressibles et modèle semi-géostrophique en météorologie.* Thèse de 3eme cycle, Université de Paris IX-Dauphine, 1992.
2. J.-D. BENAMOU and Y. BRENIER, *Weak existence for the semi-geostrophic equations formulated as a coupled Monge-Ampère/transport problem. To appear in* SIAP.
3. Y. BRENIER, *Polar factorisation and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math. 44, (1991) pp. 375-417.
4. G.R. BURTON and R.J. DOUGLAS, *Rearrangements and polar factorisation of countably degenerate functions. To appear in* Proc. Roy. Soc. Edin.
5. L.A. CAFFARELLI, *Boundary regularity of maps with convex potentials-II.* Ann. Math. 144, (1996) pp. 453-496.
6. M.J.P. CULLEN, R.J. DOUGLAS, and I. ROULSTONE, *Generalised semigeostrophic theory on a sphere. In preparation.*
7. M.J.P. CULLEN, J. NORBURY, and R.J. PURSER, *Generalised lagrangian solutions for atmospheric and oceanic flows.* Siam J. Appl. Math. 51, (1991) pp. 20-31.
8. R.J. DiPERNA and P.L. LIONS, *Ordinary differential equations, transport theory and Sobolev spaces.* Invent. Math. 98, (1989) pp. 511-547.
9. R.J. DOUGLAS, *Rearrangements of vector valued functions, with application to atmospheric and oceanic flow. To appear in* Siam J. on Math. Anal.
10. W. GANGBO and R.J. MCCANN, *The geometry of optimal transportation.* Acta Math. 177, (1996) pp. 113-161.
11. P. GERARD, *Resultats recents sur les fluides parfaits incompressibles bidimensionnels.* Seminaire Bourbaki 757, (1992) pp. 411-444.
12. J.R. HOLTON, *An introduction to dynamic meteorology.* Academic Press, 1992.
13. B.J. HOSKINS, *The geostrophic momentum approximation and the semigeostrophic equations.* J. Atmospheric Sci. 32, (1975) pp. 233-242.
14. J. JOST, *Riemannian Geometry and Geometric Analysis.* Springer–Verlag, Berlin, Heidelberg 1995.

15. M.H. MAWSON, *A shallow water semi–geostrophic model on a sphere.* Quart. J. Roy. Meteorol. Soc. 122, (1996) pp. 267-290.
16. R.J. MCCANN, *Existence and uniqueness of monotone measure preserving maps.* Duke Math. J. 80, (1995) pp. 309-323.
17. R.J. MCCANN, *Private Communication.*
18. R.J. PURSER and M.J.P. CULLEN, *A duality principle in semigeostrophic theory.* J. Atmos. Sci. 44, (1987) pp. 3449-3468.
19. G.J. SHUTTS and M.J.P. CULLEN, *Parcel stability and its relation to semigeostrophic theory.* J. Atmos. Sci. 46, (1987) pp. 2684-2697.